The method of reflective equilibrium and intuitions

Langkau, Julia

Abstract: Reflective equilibrium has been considered a paradigm method involving intuitions. Some philosophers have recently claimed that it is trivial and can even accommodate the sort of scepticism about the reliability of intuitions advocated by experimental philosophers. I discuss several ways in which reflective equilibrium could be thought of as trivial and argue that it is inconsistent with scepticism about the reliability of intuitions.

Was dürfen wir glauben?
Was sollen wir tun?

Sektionsbeiträge des achten internationalen Kongresses der *Gesellschaft für Analytische Philosophie* e.V.

Herausgegeben von
Miguel Hoeltje, Thomas Spitzley und Wolfgang Spohn

gap
gesellschaft für analytische philosophie

Was dürfen wir glauben? Was sollen wir tun?
Sektionsbeiträge des achten internationalen Kongresses der
Gesellschaft für Analytische Philosophie e.V.

Herausgegeben von
Miguel Hoeltje, Thomas Spitzley und Wolfgang Spohn

Online-Veröffentlichung der
Universität Duisburg-Essen (DuEPublico) 2013
ISBN 978-3-00-042332-1

Vorwort

Vom 17.–20.9. 2012 fand unter dem Titel „Was dürfen wir glauben? Was sollen wir tun?“ und unter der Schirmherrschaft von Frau Ministerin Prof. Schavan (Deutschland), Herrn Minister Prof. Töchterle (Österreich) und Herrn Staatssekretär Dr. Dell'Ambrogio (Schweiz) in Konstanz der achte internationale Kongress der Gesellschaft für Analytische Philosophie statt. Neben rund 35 eingeladenen Sprecherinnen und Sprechern, die in einer Reihe von Hauptvorträgen und Kolloquien zu Wort kamen, gab es in den acht thematischen Sektionen des Kongresses insgesamt mehr als 230 Vorträge und Poster-Präsentationen. Mit rund 450 Einreichungen für Sektionsbeiträge war die Beteiligung außergewöhnlich hoch. Der vorliegende Sammelband umfasst nun 61 auf solchen Sektionsbeiträgen basierende Artikel.

Ein so großer Kongress hätte nicht ohne die Beteiligung und Mithilfe vieler Menschen erfolgreich stattfinden können. Neben den Mitgliedern des GAP-Vorstandes und des GAP.8-Programmkomitees, die im Vorfeld die Planung übernommen hatten, sind hier die Mitglieder der Jurys für den Wolfgang-Stegmüller- sowie den Ontos-Preis zu nennen. Ebenfalls ist den Gutachterinnen und Gutachtern zu danken, die sich der schwierigen Aufgabe angenommen hatten, aus der übergroßen Zahl der Einreichungen die Sektionsbeiträge auszuwählen. Vor Ort in Konstanz haben ganz besonders die Kongressassistentin Gabriele Hahn und ihr Team (Wolfgang Egner, Sandra Vatter u.v.m) für eine hervorragende Planung und einen reibungslosen Ablauf der Konferenz gesorgt. Ohne die Unterstützung der DFG sowie der Universität Konstanz wäre der Kongress so nicht möglich gewesen – auch ihnen gebührt unser Dank. Ferner ist den Sponsoren von GAP.8 zu danken: dem Schweizer Staatssekretariat für Bildung und Forschung, den Fischer-Werken GmbH & Co KG, der Sparkasse Konstanz, den Verlagen de Gruyter, Klostermann, Meiner, Mentis, Ontos, Oxford University Press, Reclam, Springer, Suhrkamp und Synchron Publishers, dem philosophie-Magazin, und schließlich der Stiftung Wissenschaft und Gesellschaft, der Universitätsgesellschaft und dem Verein der Ehemaligen der Universität Konstanz.

An der organisatorischen Vorbereitung und Erstellung dieses Sammelbandes von Sektionsbeiträgen hatte niemand so großen Anteil wie unsere studentische Hilfskraft Katharina Lührmann – für die hervorragende Arbeit möchten wir schließlich ihr ganz herzlich danken!

Die Herausgeber

Miguel Hoeltje, Thomas Spitzley, Wolfgang Spohn

Inhalt

The Method of Reflective Equilibrium and Intuitions

Julia Langkau

Reflective equilibrium has been considered a paradigm method involving intuitions. Some philosophers have recently claimed that it is trivial and can even accommodate the sort of scepticism about the reliability of intuitions advocated by experimental philosophers. I discuss several ways in which reflective equilibrium could be thought of as trivial and argue that it is inconsistent with scepticism about the reliability of intuitions.

1. Introduction

Reflective equilibrium has been considered a paradigm philosophical method involving intuitions. It has been extensively discussed in normative ethics and political philosophy. The key idea as introduced by John Rawls (1971, 1974/1975) for moral and political philosophy is that we test our moral judgments (or intuitions) and moral principles against each other and revise and refine both when they are inconsistent.[1]

Without specifying what exactly they mean by it, philosophers in all areas of research frequently use the term 'reflective equilibrium' when they mention the methods and aims of their inquiry. It has been suggested that 'reflective equilibrium' is nothing more than a metaphor for the rational performance of philosophy: for taking into account all relevant information available and for working out the most plausible, coherent, and comprehensive theory of the subject matter under investigation.[2] Michael DePaul (2011) and Peter Singer (2005) suggest that it could even be compatible with scepticism about the reliability of intuitions, i.e., with a view according to which we ought not take our intuitions into account.

I am interested in the question whether the method of reflective equilibrium (MRE in what follows) is as trivial as some philosophers think or whether it gives us helpful methodological advice. I first present MRE as it has been discussed in moral philosophy (section 2), apply it to a case in epistemology (section 3), and specify what philosophers mean when they say that MRE is trivial (section 4). I then argue that the sceptical view according to which intuitions ought not be taken into account is not compatible with MRE (section 5).

2. The Method of Reflective Equilibrium

Two versions of MRE have been distinguished, *narrow* and *wide* MRE.[3] According to narrow MRE, we take (a) a set of considered moral judgments (i.e., judgments made or intuitions had in certain circumstances conducive to the truth of their content) held by a particular person

[1] Even though the term 'reflective equilibrium' was introduced by Rawls, Nelson Goodman was the first to discuss the method behind the name in 'The New Riddle of Induction' (originally 1953) in his *Fact, Fiction, and Forecast* (1955).

[2] E.g., DePaul (1998, 2011), Foley (1993), Singer (2005), and Williamson (2007).

[3] Daniels finds this distinction implicit in Rawls (1971) and explicit in Rawls (1974/1975).

and (b) a set of general moral principles and produce a coherent theory by adjusting either (a) or (b), or both. In wide MRE, we extend the area of considered judgments and principles we take into consideration to reach coherence among our *widest set of beliefs*. The following is the general idea which lies behind wide MRE, according to Norman Daniels:

> The method of wide reflective equilibrium is an attempt to produce coherence in an ordered triple of sets of beliefs held by a particular person, namely, (a) a set of considered moral judgments, (b) a set of moral principles, and (c) a set of relevant background theories. We begin by collecting the person's initial moral judgments and filter them to include only those of which he is relatively confident and which have been made under conditions conducive to avoiding errors of judgment. For example, the person is calm and has adequate information about cases being judged. We then propose alternative sets of moral principles that have varying degrees of 'fit' with the moral judgments. We do not simply settle for the best fit of principles with judgments, however, which would give us only a narrow equilibrium. Instead, we advance philosophical arguments intended to bring out the relative strengths and weaknesses of the alternative sets of principles (or competing moral conceptions). These arguments can be construed as inferences from some set of relevant background theories […] (1979: 258)

Rawls, Daniels, and other proponents of MRE think that the inclusion of alternative sets of principles and background theories is essential to a method in moral and political philosophy.[4] My focus lies on the question of how MRE applies to areas such as epistemology, metaphysics, or philosophy of language, where the subject matter under investigation does not consist of moral or similar norms. It seems that background theories play a role when we solve conflicts between our intuitions and beliefs in these areas as well. Narrow MRE will therefore not be discussed any further, and I will use 'MRE' for wide MRE in what follows.

Daniels' (1980, 1996, 2003) formulation of MRE can be given in several steps. Following the literature, I take it that 'initial moral judgments' and 'considered moral judgments' can be replaced by 'intuitions', and I will take theories to be the counterparts of moral principles in non-moral philosophy.

> Step 1. *The relevant intuitions*. Amongst all our intuitions, we choose a set of intuitions of which we are relatively confident, and which we had under ideal conditions, such as having adequate information about the subject matter under investigation and being in a state of mind that is not conducive to error.
>
> Step 2. *The best-fitting theory*. We determine a theory that fits best with the set of intuitions, i.e., that is directly supported by the set of intuitions.
>
> Step 3. *Alternative theories*. We determine alternative theories that are not directly supported by the set of intuitions, but do concern the subject matter under investigation. In the case of a counterexample to an accepted philosophical theory, our currently accepted theory will be amongst these theories.
>
> Step 4. *The relevant background theories*. We look for empirical and philosophical background theories that deliver arguments for or against the competing theories.
>
> Step 5. *Restoring coherence*. We use arguments from our background theories and our intuitions to figure out the best and most coherent theory by either disregarding our intuitions, adjusting our accepted theory, or adjusting our background theories, or all three.

---

[4] See, e.g., Rawls (1971: 49). Goodman (1955), in contrast, defends narrow MRE for the justification of inductive and deductive forms of reasoning.

3. Reflective Equilibrium and Counterexamples

Whenever some philosopher comes up with an intuitive counterexample to an accepted philosophical theory, we confront the task of resolving the resultant inconsistency in one or the other way. MRE could be the method we ought to apply to regain consistency. I will go through the steps of MRE as presented above and apply them to a well-known case of inconsistency in epistemology: the JTB theory of knowledge and the Gettier intuitions.

Step 1. Let us imagine an epistemologist Sophie who lives in the year 1963. Sophie has been thinking about knowledge for a long time, and one day she gets to read Edmund Gettier's paper 'Is Justified True Belief Knowledge?'. Here is one of Gettier's cases (for present purposes, Gettier's second case is sufficiently similar that we do not need to describe it as well):

> Suppose that while Smith has strong evidence that his friend Jones owns a Ford, he has no idea where his friend Brown is. Smith randomly selects three place-names and constructs the following three propositions: either Jones owns a Ford or Brown is in Boston; either Jones owns a Ford or Brown is in Barcelona; either Jones owns a Ford or Brown is in Brest-Litovsk. Even though Smith has no idea where Brown is, he is justified in believing each of these three propositions, because he has correctly inferred them from a proposition for which he has strong evidence, namely that Jones owns a Ford. However, Jones does in fact not own a Ford, but is driving a rented car. Unknown to Smith, Brown happens to be in Barcelona. It seems that Smith does not know that either Jones owns a Ford or Brown is in Barcelona, even though it is true, Smith believes that it is true, and Smith is justified in believing that it is true.[5]

Sophie agrees with the description of the case. She has the intuition that Smith does not know that either Jones owns a Ford or Brown is in Barcelona. However, the theory of knowledge that Sophie holds, the JTB theory, predicts that Smith knows that either Jones owns a Ford or Brown is in Barcelona. Hence, the content of Sophie's intuition is clearly inconsistent with what follows from the JTB theory. Suppose that Sophie notices the inconsistency and strives to resolve it. Suppose furthermore that Sophie's intuition meets the relevant criteria: Sophie is relatively confident of the intuition, and she had it under ideal conditions. Her intuition therefore qualifies for MRE (in Daniels' terms, the intuition is a 'considered judgment' as opposed to an 'initial judgment').

Step 2 and step 3. Sophie finds the JTB theory supported by a set of her intuitions about cases where justified true belief seemingly is knowledge. For instance, Sophie has the following intuition: if she hears her cat meowing in the kitchen she is justified in believing that her cat is in the kitchen, and if it is also true that her cat is in the kitchen, then she knows that her cat is in the kitchen. Her intuitions in the Gettier Cases are not part of this set of intuitions, and Sophie now has to determine the theory that fits her Gettier intuitions best as well as alternative theories that compete with the best-fitting theory.

Sophie reads Michael Clark's response to Gettier, 'Knowledge and Grounds: A comment on Mr. Gettier's Paper'. She spends a long time thinking about the cases and about Clark's no-false-lemma reply to Gettier. In both Gettier Cases, Smith's reasoning is based on a false premise. In the case quoted above, Smith gains his justified true belief by reasoning from the justified false belief that Jones owns a Ford. The case Clark gives in reply has it that knowledge is justified true belief where the justification is not based on a false assumption. Sophie also comes up with her own theories, of which the first is a defeasibility analysis of knowledge. The defeasibility account has it that knowledge is justified true belief absent a

[5] Gettier (1963: 14-15).

defeater.[6] The second theory is a causal theory, on which the belief that P is knowledge only if it is appropriately causally connected to the fact that P.[7] The third theory Sophie comes up with involves a simple reliability condition. According to such an account, S knows that P if and only if S's belief that P is true and justified, where S's belief that P is justified if and only if the belief that P was produced by a reliable cognitive process.[8] Sophie then decides that Clark's adjusted JTB theory fits best with her intuitions in the Gettier Cases (step 2). The JTB theory, the causal theory, the defeasibility account, and the reliability account are alternative theories Sophie has to consider (step 3).

Step 4. Sophie needs to think about her background theories which will help her to decide between the JTB theory, Clark's account, and the alternative theories. Which theory of knowledge Sophie chooses certainly depends on what she thinks justification is. Theories of justification are relevant yet do not directly concern the question of what knowledge is. To keep Sophie's case simple, let us suppose that she has externalist views about justification, which means that she will not consider internalist theories of knowledge. Metaphysical accounts of causation might also influence Sophie's choice. Since metaphysics of causation suggests that abstract and future facts cannot be causes, this counts against the causal theory of knowledge. Other background theories will probably influence Sophie's preferences tacitly.

Step 5. What Sophie has done so far should help her to remove the inconsistency between the content of her intuition that Smith does not know the relevant proposition and her accepted theory of knowledge, the JTB theory.

Sophie thinks that Clark's account is the one that fits her Gettier intuitions best, but she now wonders whether it covers other cases as well and tries to come up with a counterexample. After a while of thinking, she comes up with a case where the justification does not rest on a false assumption, similar to Keith Lehrer's (1965) Nogot case. In Lehrer's case, Nogot in S's office has given S evidence that he, Nogot, owns a Ford. S directly moves from his evidence to the conclusion that someone in S's office owns a Ford, without arguing via the assumption that Nogot owns a Ford. Nogot does not own a Ford, however, someone else in the office, Havit, owns one, which makes S's belief true. Let us suppose that Sophie's case is very similar to this case in that the subject's reasoning somehow does not rest on a false assumption. Sophie comes to the conclusion that Clark's analysis is not correct, because it cannot account for some cases that are very similar to the original Gettier Cases.

Let us say that the causal theory is not consistent with Sophie's background theory on causality. The defeasibility analysis she came up with covers more cases than Clark's analysis, but Sophie thinks it is also extremely complicated.

Sophie thinks about the advantages and disadvantages of the JTB theory, the theory she thinks fits her Gettier intuitions best, and the alternative theories. She weighs them against each other and, given her views on justification and causation which she does not want to give up, decides that a reliabilist account of knowledge is the best choice. While Sophie's case is fictional and could have been told differently, it obviously roughly corresponds to how parts of the debate about the Gettier Cases were conducted in the literature, and some philosophers made a similar decision as Sophie.

---

[6] See Lehrer & Paxon: 'A defeasibility condition requires that there is no other true statement, d, such that the conjunction of S's present evidence for p with d would fail to make S justified in believing p.' (1969, p. 230).

[7] See, e.g., Goldman (1967).

[8] See, e.g., Goldman (1979).

4. Reflective Equilibrium and the Triviality Charge

Let us now look at two obvious ways in which MRE could be interpreted as non-trivial or misguided. First, some philosophers understand MRE as a theory of justification (e.g., Rawls (1951), Daniels (1979, 1980, 1996), Elgin (1996), Stich (1988), and Goodman (1955)). According to MRE as a theory of justification, the beliefs we reach as a result of applying MRE to a certain topic are thereby justified. For instance, if Sophie comes to the conclusion that Smith in the Gettier Case does not know that either Jones owns a Ford or Brown is in Barcelona, then Sophie's belief that Smith does not have knowledge of this proposition is justified if it is in reflective equilibrium, i.e., if it coheres with all her other beliefs about knowledge as a result of having appropriately followed steps 1 to 5. As such, MRE is most naturally understood as a coherentist theory of justification. According to coherentist theories of justification, our beliefs are justified through their relation to other beliefs.[9] Coherentist theories of justification are controversial, especially for the justification of our beliefs in areas other than moral or political philosophy. Clearly philosophers such as Timothy Williamson (2007) who claim that reflective equilibrium is trivial do not have reflective equilibrium as an account of justification in mind.

What else could MRE be if not a theory of justification? Besides merely aiming for true beliefs, we aim to build our theories or revise our beliefs in a rational manner. Ideally, we want a methodology in the sense of rules that guide us in the process of building a theory or revising our beliefs. MRE could simply be understood as a method of rational belief revision that does not carry any commitment with regard to what exactly we gain when we apply MRE.[10] As a method of rational belief revision, MRE is neutral with respect to epistemic theories such as foundationism or coherentism, and it is neutral with respect to epistemic externalism or internalism as well. Distinguishing a method of rational belief revision from the question whether our beliefs are justified or true accounts for the strong intuition that our opponents sometimes are as rational in holding the beliefs they hold as we ourselves are, even if either our opponents or we ourselves have unjustified and false beliefs.[11] Supposing that we are not in a sceptical situation, MRE as a method of rational belief revision could nevertheless be conducive to the justification of our beliefs. It is prima facie plausible that if a subject S is not deceived by a Cartesian demon and starts out with beliefs that are rational for her to hold, and if she revises her beliefs in a rational way to reach a theory T, then it is more likely that T is justified than that T is not justified. In different terms, if S is not victim of a Cartesian demon and starts out with beliefs that are rational for her to hold, and if she revises her beliefs in a rational way to reach a theory T, then this is at least some evidence for the truth of T. This does not mean that coherence of T is the only or even the best evidence for the truth of T. In what follows, I will take MRE to be a method of rational belief revision.

Here is a second way in which MRE could be interpreted as non-trivial: it could be understood as misguided because it idealizes our practice of revising beliefs. DePaul (2011) claims that this is the most serious problem MRE confronts, and that MRE can be understood

---

[9] See Daniels (2011).

[10] Philosophers have discussed several ways in which MRE could be a useful method in moral philosophy. Geoffrey Sayre-McCord (1996) mentions three approaches that do not involve the justification of beliefs. According to the first, MRE is a heuristic method, i.e., MRE is useful to discover the fundamental truths of morality, but it does not justify the beliefs reached through its application. According to a second approach, there is a moral obligation to act only upon moral principles that are in reflective equilibrium with all our other beliefs. According to a third and pragmatic view, it is in some sense useful to act upon a principle which is in reflective equilibrium with all other beliefs we hold. See also Kappel (2006), who is pessimistic concerning the role of MRE in epistemic justification, but mentions that a pragmatic or otherwise not truth-related account of MRE might be defendable.

[11] See also Kelly & McGrath (2010), who think that MRE ought to lead to beliefs that are rational for us to hold.

as idealizing our practice with regard to two aspects. First, the order in which we ought to proceed according to MRE does not correspond with what we in fact do: we hardly ever start out with first determining the relevant intuitions (step 1), figuring out a set of theories that matches (step 2), then taking alternative theories into consideration (step 3), and finally using our background theories (step 4) to build the most comprehensive and coherent theory on the subject matter (step 5). We rather 'naturally bring all kinds of considerations [...] into play helter skelter as they occur to us'[12]. Second, the quantity of intuitions and theories we take into account is limited. It is simply not possible for us to take all relevant alternative theories into account (step 3), and it is not possible to test our intuitions concerning all relevant aspects (step 1), because 'one would need to reflect upon and form [intuitions] about far too many kinds of hypothetical cases'[13]. Looking at Sophie's case, one might think this is indeed a problem. Surely the way Sophie applies MRE to the Gettier Cases is an idealization. No single philosopher did apply or could have applied MRE in the way Sophie does: Sophie starts out with step 1 and then goes through steps 2 to 5. She also takes far more theories into consideration in steps 2 and 3 before she moves on to the final step 5 than philosophers did in 1963.

There are two ways to reply to the idealization objection. First, we might say that MRE has to be understood as making claims about what we should ideally do, not as telling us what we ought to do given our temporal and intellectual constraints. MRE says that we should consider all possible alternatives, but this does not mean that it is not rational for us to stop at a certain point to settle for a coherent theory. Let us look at Sophie again. Even though she considers many alternative theories, one could still criticize her for not doing everything she ought to do in order to appropriately follow the advice MRE gives. Maybe Sophie should not have made a decision as to which theory to endorse, since the debate over the Gettier Cases is ongoing and philosophers are still coming up with theories to cover our intuitions in the Gettier Cases and numerous variations of the Gettier Cases. Ideally, Sophie would even be much smarter. However, just as it seems rational for Sophie to stop considering alternative thought experiments and alternative theories, other philosophers could be rational relative to their limits.

It is unproblematic to stop considering alternative theories because MRE can be applied to a minimal set of intuitions and theories. MRE can be re-applied whenever someone comes up with a new theory or a new thought experiment, and the accepted theory can be revised again. This is a reply to both aspects of the idealization objection, the worry that we do not always go through the steps in the right order and the worry that we cannot consider all relevant theories and intuitions. To engage in MRE could simply mean to engage in an ongoing series of applications of MRE. As Rawls claims: once a subject has reached a coherent theory, "[...] this equilibrium is not necessarily stable. It is liable to be upset by further examination [...] and by particular cases which may lead us to revise our judgments [...]"[14].

In fact, as I presented Sophie's case, Sophie comes up with a new thought experiment in step 5, namely with a case that is similar to Lehrer's Nogot case. The reason why I let Sophie come up with this case in step 5 is that she compares different theories in step 5, and hence she thinks about whether Clark's account can accommodate as true all or enough intuitions only in step 5. Strictly speaking, Sophie goes back to step 1. However, it seems that this is not a problem for MRE: the order in which we follow the steps does not seem to be crucial.

A second way to reply to the idealization objection is to say that MRE ought to be applied by a group of researchers instead of a single philosopher. Whereas Rawls, Daniels, and DePaul think that MRE is to be pursued by an individual researcher, Goodman (1955) defends such a

[12] DePaul (2011: xcii).
[13] DePaul (2011: xcii).
[14] Rawls (1971: 18).

collective MRE. The view is that as individual philosophers, we work on different ends: we discover inconsistencies, we develop theories that cover our intuitions best, and we develop background theories. We do all this in much detail, which might require a whole career or life-time. As a community of researchers, we might eventually reach the aim of step 5: one single theory wins. Maybe Sophie might better be understood as representing a group of researchers, not an individual philosopher.

DePaul argues that MRE cannot possibly be understood as the description of what we are supposed to do as a group of philosophers:

> Because of the way revisions are determined, [reflective equilibrium] must be a first-person inquiry. Propositions do not seem true to a group of people except in the derivative sense that they seem true to each member of the group. Any disagreement within a group and there will be nothing to determine how conflicts are to be resolved, and hence, nothing to determine the group's state of [reflective equilibrium]. Moral inquiry can be a joint endeavor according to [reflective equilibrium] only insofar as we agree or insofar as one person is willing to assist another in her individual attempt to bring her beliefs into equilibrium by doing such things as pointing out potential conflicts in her beliefs, presenting examples that might elicit interesting intuitions or proposing theories that might account for her [intuitions]. Alternatively, one might approach some other person as a subject, taking that person's beliefs and seemings as data and attempting to work out what that person's state of [reflective equilibrium] would be. (2011: 1xxxi)

DePaul thinks that disagreement in intuitions makes a collective MRE impossible. However, disagreement might simply show that we have not yet reached the final state of reflective equilibrium. There surely is a lot of disagreement in philosophy, and this disagreement concerns our theories as much as our intuitions. We could nevertheless all be concerned with the same project, namely with finding the best theory of a certain subject matter, and it is still possible that in the end one theory will win in the sense that it will be accepted by everyone.

While I do not think that disagreement makes collective MRE impossible, I agree with DePaul that MRE should be understood as a method an individual philosopher ought to pursue. The reason is that if MRE is what a group of philosophers ought to engage in, it is unclear what the advice for the individual philosopher would be and how we would access whether an individual philosopher is revising her beliefs in a rational way. This, however, is what we are interested in when we talk about a method of rational belief revision. We want such a method to give us advice on how to revise our beliefs as individual philosophers and we want to be able to decide whether an individual philosopher is revising her beliefs rationally. Even if other people's intuitions matter, it does not follow that MRE is a collective enterprise. It is plausible to think that we do not only take our own intuitions into consideration but rather rely on other philosophers with respect to intuitions as much as we do with respect to theories. Endorsing MRE as a method for an individual philosopher does not mean that we cannot rely on work other researchers have done, on their intuitions or on the theories they developed.

If we do not understand MRE as a theory of justification and if we think it does not have to be viewed as an idealization of our practice, the concern is that step 1 to 4 of MRE are trivial and step 5 gives us only very general instructions. According to Rawls, we have to go "back and forth", sometimes to adjust the principles to our judgments, sometimes to conform the judgments to our principles.[15] Similarly vaguely, Daniels claims that we are "expected to revise our beliefs at all levels as we work back and forth among them and subject them to various criticisms"[16]. It seems that MRE collapses into a trivial and uncontroversial claim

[15] Rawls (1971: 20).
[16] Daniels (2003).

about philosophical methodology, as Williamson expresses in the following passage about MRE:

> The question is not whether philosophers engage in the mutual adjustment of general theory and judgments about specific cases—they manifestly do—but whether such descriptions of it are sufficiently informative for epistemological purposes. (2007: 244)

Similarly, Foley thinks that MRE is too general to be useful:

> The problem with this recommendation is familiar. It is not so much mistaken as unhelpful. [...] It tells you essentially this: take into account all the data that you think to be relevant and then reflect on the data, solving conflicts in the way that you judge best. On the other hand, it does not tell you what kinds of data are relevant, nor does it tell you what is the best way to resolve conflicts among the data. It leaves you to muck about on these questions as best you can. (1993: 128)

DePaul (1998, 2011) goes so far as to argue that MRE is the only rational method in philosophy. In a nutshell, DePaul's version of MRE says that we should reflect upon the logical and evidential relations between all our relevant beliefs, and that we should resolve conflicts which might emerge during this process in the best possible way we can figure out. DePaul then argues that it is difficult to think of a rational alternative to MRE thus construed, and that an opponent of MRE would have to make one of the following claims: (A) we should abandon reflection altogether; (B) our method should direct the inquirer to reflect, but to do so incompletely, i.e., to leave certain beliefs, principles, or theories out of account; (C) our method should not allow the results of the inquirer's reflections to determine what the inquirer goes on to believe.[17] Unsurprisingly, DePaul concludes from his discussion of (A), (B), and (C) that any method endorsing at least one of these claims would be irrational.

One way of deciding whether MRE is trivial is to see whether it is compatible with different views concerning the role of intuitions in philosophy. If it is not, then it seems that MRE is not trivial. In the next section, I argue that MRE is not compatible with the sort of scepticism about the reliability of intuitions advocated by experimental philosophers.

5. Reflective Equilibrium and Scepticism

Some experimental philosophers draw a radical conclusion from their studies: using intuitions as evidence for or against philosophical theories is an unreliable method which should not be pursued.[18] Some philosophers have claimed that MRE can accommodate scepticism about the reliability of intuitions as advocated by experimental philosophers. I first look at a view according to which sceptical considerations are not supposed to be part of MRE because meta-theories in general are not supposed to be part of MRE (section 5.1). I then look at Singer's scepticism about intuitions (5.2) and at DePaul's view according to which scepticism is compatible with MRE because it could come as a result of applying MRE that we disregard every single intuition (section 5.3). I conclude that MRE is not compatible with scepticism about the reliability of intuitions.

5.1 MRE and Meta-Theories

One way to argue that scepticism about the reliability of intuitions is not compatible with MRE is to exclude meta-considerations from the beginning. DePaul mentions that according to Rawls' original account, meta-considerations and arguments from philosophy of language and metaphysics that have been used for or against metaethical theories such as moral

[17] DePaul (1998: 301).

[18] E.g., Alexander & Weinberg (2007), and Weinberg (2007).

relativism, moral realism, or noncognitivism are not supposed to be considered in MRE. The reason is that while MRE is supposed to help us decide between different moral theories, meta-theories do not bear directly on the moral subject matter under consideration.[19] The same would apply to contemporary empirical research on the reliability of our intuitions, according to DePaul:

> Efforts to use results from psychology or neuroscience or evolutionary theory to question the reliability of some or all of our [intuitions] are now extremely prominent [...] As Rawls conceived of [wide reflective equilibrium], these background theories would not be part of the equilibrium. Because they provide premises for a broad skepticism regarding morality, they would not serve as premises of arguments for, or against, any particular [moral theory]. (2011: lxxxix)

Exactly the same would be true for results from psychology or experimental philosophy in epistemology. Take Sophie and her epistemic intuitions. Sophie is interested to know whether Smith has knowledge in the Gettier Case, and more generally what knowledge is. Questions concerning the relevance of empirical research on intuitions do not bear directly on the question whether Smith has knowledge of the relevant proposition or not, so it seems not to help Sophie to answer her question.

DePaul mentions that Rawls' main reason to exclude meta-theories from MRE is that he is interested in MRE as a method to detect our moral sensibilities rather than to detect the truth about moral issues. In order to determine which moral theory captures our moral sensibilities best, meta-theories are obviously not relevant. To compare it with a simple case: if I am interested to find out about my food preferences and notice that I do not like broccoli, the fact that broccoli is healthy and it would be much better for me to like broccoli is irrelevant to my concern.

However, if we are ultimately interested in the truth of our theories rather than merely in the systematization of our beliefs and intuitions, the question of whether our intuitions are reliable is highly relevant. Hence, if MRE told us to ignore doubts as to their reliability, it could not be a method of rational belief revision. Excluding meta-considerations such as scepticism about the reliability of intuitions from the beginning is thus not an option.

5.2 MRE without Intuitions

In line with some experimental philosophers, Singer (2005) draws sceptical conclusions from recent empirical studies and defends a view according to which we should not assign any plausibility to our moral intuitions. Singer objects to "any method of doing ethics that judges a normative theory either entirely, or in part, by the extent to which it matches our moral intuitions"[20].

As an example of evidence for the insignificance of moral intuitions, Singer discusses Joshua Greene et al.'s (2001) studies on intuitions people have when confronted with different versions of the Trolley Case. In one version of the Trolley Case, we are asked whether a fat man should be pushed down a bridge to stop the trolley, in which case only one person dies and five people who would otherwise be killed survive. Most of us have the intuition that pushing the fat man off the bridge would be wrong. In another version, we are asked if the driver of the trolley should side-track the trolley, which would again kill only one person instead of five people. Most of us have the intuition that the driver should side-track the trolley. Greene et al. (2001) conducted brain scans of subjects while they had intuitive reactions to cases very similar to the two versions of the Trolley Case. The results suggest that different intuitive responses have to be explained by differences in the emotional pull of

[19] DePaul (2011: lxxxix).

[20] Singer (2005: 346).

situations which involve causing someone's death in a close-up, personal way vs. causing someone's death in a way which is at a distance and less personal.[21] Singer thinks that more research is likely to show that Greene has not only explained, but rather explained away the philosophical puzzle of why our intuitions in different versions of the Trolley Case differ. Based on Greene et al.'s results and arguments from evolution[22], Singer draws the following sceptical conclusions for moral intuitions and for MRE:

> [R]ecent scientific advances in our understanding do have some normative significance, and at different levels. At the particular level of the analysis of moral problems like those posed by trolley cases, a better understanding of the nature of our intuitive responses suggests that there is no point in trying to find moral principles that justify the differing intuitions to which the various cases give rise. Very probably, there is no morally relevant distinction between the cases. At the more general level of method in ethics, this same understanding of how we make moral judgments casts serious doubt on the method of reflective equilibrium. There is little point in constructing a moral theory designed to match considered moral judgments that themselves stem from our evolved responses to the situations in which we and our ancestors lived during the period of our evolution as social mammals, primates, and finally, human beings. (2005: 348)

We are interested in the second claim on a more general level, according to which empirical studies cast doubt on MRE. Even though Singer does not think it would be a good idea, he mentions that MRE could possibly be interpreted "wide enough" to accommodate a practice where we do not take any of our intuitions into account. In that case, MRE might be compatible with the idea that intuitions should not play any role in philosophy, but it would no longer be a distinctive method for normative ethics.[23]

Scepticism about the reliability of intuitions entails *prima facie* that step 1 is misguided: we ought not take our intuitions into account. In a concrete case of a thought experiment such as the Gettier Cases, this means that a subject ought not assign any plausibility to the fact that she has an intuition that P or to the content of the intuition that P. In general terms:

> **No Plausibility**
>
> The subject assigns no plausibility to her intuition that P.

A method that rules out intuitions (either the fact that we have an intuition that P or the content of the intuition that P) as evidence from the beginning does not seem to be compatible with MRE, at least not with MRE as presented in section 2. According to MRE, we remove an inconsistency between an intuition and our accepted theory either by disregarding the intuition and revising our judgment or by adjusting our accepted theory (or our background theories, or all three). The crucial point is that two options are available with respect to the intuition that P: either to accommodate it as true or to disregard it. In principle, MRE leaves Sophie the choice to either endorse that Smith has no knowledge, or to reject it and either stick with the JTB theory or endorse a different theory. Which option she chooses depends on her background theories and on how good the alternative theories she comes up with are, but neither is ruled out from the beginning.

---

[21] Greene et. al (2001: 2106).

[22] According to Singer, "[...] the salient feature that explains our different intuitive judgments concerning the two cases is that the footbridge case is the kind of situation that was likely to arise during the eons of time over which we were evolving; whereas the standard trolley case describes a way of bringing about someone's death that has only been possible in the past century or two, a time far too short to have any impact on our inherited patterns of emotional response." (2005: 348).

[23] Singer (2005: 347).

Proponents of MRE might disagree about how exactly the process of restoring coherence in step 5 ought to take place. We said that background theories are relevant to the areas we are interested in, i.e., the relevant MRE is wide MRE. This means that we use arguments from background theories in step 5, and it also means that background theories are amongst the theories that can be adjusted. However, that there are two ways of dealing with the intuition seems to be widely shared. Here is how Ernest Sosa puts it (using 'principle' for 'theory'):

> If a conflict pits the intuitive pull of an example against the tug of a familiar principle, we seek to remove or revise one or the other, so as to remove the tension. Sometimes the particular intuition(s) will win, but sometimes the tug of the principle must prevail. (1989: 262)

Roy Sorensen writes (using 'theoretical principles' for 'theories' and 'atheoretical judgment' for 'intuition'):

> You attain reflective equilibrium when your theoretical principles cohere with your atheoretical judgments. To reach this state, you must remove conflicts between them. Sometimes the conflict is resolved by giving up the principle and sometimes by giving up the intuition. (1992: 83)

It seems that step 5 of MRE, applied to thought experiments as counterexamples to philosophical theories, entails the following methodological claim with respect to intuitions.

> **Coherence**
>
> In case of a conflict between our intuitions and our accepted theory on a certain subject matter it is sometimes permissible to disregard an intuition that P and sometimes permissible to adjust theory to accommodate as true an intuition that P.

If we endorse No Plausibility, Coherence must be false. Since according to scepticism about the reliability of intuitions, we ought not assign any plausibility to our intuitions, it is not true that we have two options to regain coherence. According to this view, thought experiments simply would not be potential counterexamples and would not have to be taken into consideration at all. Sophie, for instance, would have to disregard her intuition in the Gettier Case and stick with her accepted theory, the JTB theory of knowledge—or use other kinds of evidence to support a different theory of knowledge. According to MRE, however, we have to assign at least some initial plausibility to our intuitions. Hence, if scepticism about the reliability of intuitions entails No Plausibility, it is not compatible with MRE.

5.3 Scepticism as a Consequence of MRE

However, there might be a way to account for scepticism about the reliability of intuitions that does not entail No Plausibility. Let us look at DePaul's view. In contrast to Singer, DePaul thinks that moral intuitions play an important role, but he agrees that MRE could accommodate scepticism about the reliability of intuitions.[24] As a consequence, DePaul's characterization of MRE is indeed so general that it could be true of any kind of truth-directed activity: the essence of MRE is that it 'directs one to leave nothing out of consideration and to believe what seems likely to be true upon due consideration'.[25] It seems that MRE understood as broadly as this can no longer count as a method of rational belief revision since it does not entail any advice as to how to revise our beliefs.

DePaul thinks that scepticism about the reliability of intuitions could come as a result of applying MRE. If we take empirical research on particular intuitions such as psychological and neuroscientific evidence into account, it could turn out that we end up rejecting every

[24] DePaul (2011: c).

[25] DePaul (2011: cii).

single one of our intuitions. This seems to be what DePaul has in mind when he says the following:

> Suppose the data provided by the research seems much more likely to be true to S than any of her intuitive moral judgments, and that she follows the arguments from the data and has no doubts about them. So, S excises all normative moral beliefs from her overall system. She ends up accepting no moral theory; she makes no moral judgments. She only has beliefs about morality, e.g., that all the moral judgments she previously made were mistaken and that all the moral judgments other people make are mistaken. Does it follow that S would have abandoned the method of [reflective equilibrium]? Not at all—she would have done exactly what that method dictates. (2011: c)

Since he thinks that all moral theory in the end amounts to intuitions, DePaul concludes that if the subject rejected all her moral intuitions, she would have no moral views at all. We would not have to take this extreme stance on the role of intuitions in non-moral philosophy, but we could imagine ending up disregarding all intuitions from thought experiments, because it turned out that we have reasons to do so for every single intuition.

DePaul's idea is that since it would come as a result of applying MRE, disregarding all intuitions would be compatible with MRE. However, such a result would certainly raise serious doubts about the reliability of intuitions more generally. We would then have to adjust our meta-philosophical view: it would hardly be rational to consider further intuitions, and we would rather have to assign no initial plausibility to intuitions anymore. We would have to endorse No Plausibility, which entails the denial of Coherence (as discussed above). Hence, if MRE leads to the rejection of every single of our intuitions, it is self-defeating. However, as long as this does not lead us to disregard all our intuitions, MRE is compatible with taking empirical evidence about the reliability of particular intuitions into account.

6. Conclusion

The question I aimed to answer in this paper was whether MRE understood as a method of rational belief revision reduces to a trivial claim about philosophical methodology. I argued that contrary to what some philosophers think, MRE is not compatible with scepticism about the reliability of intuitions and is thus not trivial.

The claim that we ought to assign some plausibility to our intuitions is ambiguous. It can either mean that we take the content of the intuition that P or that we take the fact that we have an intuition that P to be evidence. In order to understand the role of intuitions in MRE, more work has to be done to specify how exactly they are supposed to come into play.

Julia Langkau

University of Zurich
julialangkau@gmail.com

References

Alexander, J., and J. Weinberg 2007: 'Analytic epistemology and experimental philosophy', *Philosophy Compass* 2, 56–80.

Daniels, N. 1979: 'Wide reflective equilibrium and theory acceptance in ethics.' *The Journal of Philosophy* 76, 5 (1979), 256–282.

— 1980: 'Reflective equilibrium and archimedean points', *Canadian Journal of Philosophy* 10, 83–103.

— 1996: *Justice and Justification: Reflective Equilibrium in Theory and Practice*. New York: Cambridge University Press.

— 2003: 'Reflective equilibrium', http://plato.stanford.edu/entries/ reflective-equilibrium/ (2003; substantive revision 2011).

DePaul, M. R. 1998: 'Why bother with reflective equilibrium?', in *Rethinking Intuition. The Psychology of Intuition and Its Role in Philosophical Inquiry,* Rowman and Littlefield Publishers, Inc., 293–309.

— 2011: 'Methodological issues: Reflective equilibrium', in *The Continuum Companion to Ethics*, London, New York: C. Miller, Ed. Continuum International Publishing Group, Ixxv–cv.

Elgin, C. Z. 1996: *Considered Judgment*. Princeton University Press.

Foley, R. 1993: *Working Without a Net: A Study of Egocentric Epistemology*. New York: Oxford University Press.

Gettier, E. L. 1963: 'Is justified true belief knowledge?', *Analysis* 23, 121-123.

Goldman, A. 1967: 'A causal theory of knowledge', *Journal of Philosophy* 64, 357-372.

— 1979: 'What is justified belief?', in *Justification and Knowledge: New Studies in Epistemology*, Boston, Dordrecht, London: Reidel, 1–23.

Goodman, N. 1955: *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.

Greene, J. D., R. B. Sommerville, L. E. Nystrom, J. M. Darley, and J. D. Cohen 2001: 'An fMRI investigation of emotional engagement in moral judgment', *Science* 293, 2105–2108.

Kappel, K. 2006: 'The meta-justification of reflective equilibrium', *Ethical Theory and Practice* 9, 131–147.

Kelly, T., and S. McGrath 2010, 'Is reflective equilibrium enough?', *Philosophical Perspectives Volume* 24, 325–359.

Lehrer, K. 1965: 'Knowledge, thruth, and evidence', *Analysis* 25, 168–175.

Lehrer, K., and T. Paxon 1969: 'Knowledge: Undefeated justified true belief', *Journal of Philosophy* 66, 225–237.

Rawls, J. 1951: 'Outline of a decision procedure for ethics', *The Philosophical Review* 60, 177–197.

— 1971: *A Theory of Justice*. Harvard: Cambridge University Press, 1971.

— 1974/1975: 'The independence of moral theory', *Proceedings and Addresses of the American Philosophical Association* XLVII, 5–22.

Sayre-McCord, G. 1996, 'Coherentist epistemology and moral theory', in *Moral Knowledge?,* Sinnott-Armstrong, W. , and M. Timmons (eds.), Oxford University Press, 137–189.

Singer, P. 2005, 'Ethics and intuitions', *The Journal of Ethics*, 331–352.

Sorensen, R. 1992: *Thought Experiments*. New York: Oxford University Press.

Sosa, E. 1989: 'Equilibrium in coherence?', in *The Current State of the Coherence Theory*, J. Bender (ed.), Kluwer: Academic Press, 242–250.

Stich, S. 1988: 'Reflective equilibrium, analytic epistemology and the problem of cognitive diversity', *Synthese* 74.

Weinberg, J. 2007: 'How to challenge intuitions empirically without risking skepticism', *Midwest Studies in Philosophy* 31, 318–343.

Williamson, T. 2007: *The Philosophy of Philosophy*. Blackwell Publishing Ltd.